DOCUMENT RESUME

ED 247 284 TM 840 462

AUTHOR	Raudenbush, Stephen W.; Bryk, Anthony S.
TITLE	Application of Empirical Bayes Estimation in Educational Research.
PUB DATE	Apr 84
NOTE	18p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Bayesian Statistics; *Estimation (Mathematics); *Mathematical Models; Research Methodology; Statistical Analysis
IDENTIFIERS	*Empirical Bayes Estimation

ABSTRACT

The purpose of this paper is to demonstrate in detail how the Empirical Bayes (EB) statistical estimation strategy can be applied to an important class of educational research contexts. EB methods are tailored specifically to the analysis of data with a hierarchical structure. For instance, investigators may be interested in discovering how effects within schools (e.g., the relationship between student social class and achievement) vary as a function of differences between schools (e.g., policies and practices). Similarly, meta-analysts often wish to find out how differences between experimental and control groups within studies vary as a function of differences between studies (e.g., how treatments are implemented). Developmental psychologists care about how children's intellectual growth rates vary as a function of different pre-school experiences. In each case parameters at one level (within schools, within studies, and within children) vary as a function of parameters at another level (between schools, between studies, between children). This paper explains how the EB strategy works when the central goal of an investigation is to estimate the second level parameters (i.e., the between-group parameters), and an important ancillary goal is to assess the adequacy of a hierarchical linear model for fitting such hierarchical data. (Author)

APPLICATION OF EMPIRICAL BAYES ESTIMATION IN EDUCATIONAL RESEARCH

Stephen W. Raudenbush
Michigan State University

Anthony S. Bryk
Harvard University

Paper to be presented at the annual meeting of the American Educational Research Association

New Orleans, 1984

The purpose of this paper is to demonstrate in detail how the "empirical Bayes" or "James-Stein" statistical estimation strategy (Lindley and Smith, 1972; Strenio, 1981; Strenio, Bryk, and Weisberg, 1983; Raudenbush, 1984: note 1) can be applied to an important class of educational research contexts. Empirical Bayes ("EB" for short) methods are tailored specifically to the analysis of data with a hierarchcial structure. For instance, investigators may be interested in discovering how effects within schools (e.g., the relationship between student social class and achievement) vary as a function of differences between schools (e.g., policies and practices). Similarly, meta-analysts often wish to find out how differences between experimental and control groups within studies vary as a function of differences between studies (e.g., how treatments are implemented). Developmental psychologists care about how children's intellectual growth rates vary as a function of different pre-school experiences. In each case parameters at one level (within-schools, within studies, and within children) vary as a function of parameters at another level (between schools, between studies, between children). This paper explains how the EB strategy works when the central goal of an investigation is to estimate the second level paramters (i.e., the between-group paramters), and an important ancillary goal is to assess the adequacy of a hierarchcial linear model for fitting such hierarchical data.

The purpose of this paper is to demonstrate in detail how the "empirical Bayes" or "James-Stein" statistical estimation strategy (Lindley and Smith, 1972; Strenio, 1981; Strenio, Bryk, and Weisberg, 1983; Raudenbush, 1984: note 1) can be applied to an important class of educational research contexts. Empirical Bayes ("EB" for short) methods are tailored specifically to the analysis of data with a hierarchcial structure. For instance, investigators may be interested in discovering how effects within schools (e.g., the relationship between student social class and achievement) vary as a function of differences between schools (e.g., policies and practices). Similarly, meta-analysts often wish to find out how differences between experimental and control groups within studies vary as a function of differences between studies (e.g., how treatments are implemented). Developmental psychologists care about how children's intellectual growth rates vary as a function of different pre-school experiences. In each case parameters at one level (within-schools, within studies, and within children) vary as a function of parameters at another level (between schools, between studies, between children). This paper explains how the EB strategy works when the central goal of an investigation is to estimate the second level paramters (i.e., the between-group paramters), and an important ancillary goal is to assess the adequacy of a hierarchcial linear model for fitting such hierarchical data.

We proceed by explaining the empirical Bayes concept concretely in the context of school effectiveness research. A reanalysis of the High School and Beyond data (Coleman, Hoffer, and Kilgore, 1982) demonstrates the use of the method in simultaneously estimating between- and within-school effects.

Methodological Difficulties in School Effects Research

A central difficulty in quantitative investigation of school effectiveness is the multi-level character of the data. There are differences between schools and differences among the students within each school. To be valid, statistical analyses must account simultaneously for effects at both levels. The most promising current approaches (Cooley, Bond, and Mao, 1981; Burstein and Miller, 1978) first estimate relationships between student background characteristics and student outcomes at each school; these relationships, as specified by regression coefficients, then serve as outcome variables for an assessment of the importance of school policies and practices.

However, several major difficulties typically plague the "slopes as outcomes" analysis:

1. Within-school regression coefficients are typically estimated unreliably: far more unreliably than, say, school means. Such sampling variance is especially acute when sample sizes within schools are small and/or several regression coefficients per school are estimated. This unreliability of slopes as outcomes means that even if some schools are more effective than others in, say, minimizing the effect of student social class on achievement, the analysis may fail to detect this effect or seriously underestimate its importance.

2. "Slopes-as-outcomes" research very often uses a single regression slope per school as an outcome variable. For instance, one might use the SES/achievement slope as an outcome. Yet it will typically be necessary to adjust for other variables which might confound the SES/achievement relationship.

However, to include multiple predictors within schools requires a statistical model which takes into account the interdependence of these coefficients.

The EB approach helps the investigator overcome these difficulties in many instances. First, it enables the analyst to distiguish between two sources of variation among the estimates of the within-school slopes: the variation of the paramters themselves, which actually constitute the object of the investigation; and the sampling variance of the estimates. Second, the EB approach takes into account the unequal precision of the within-school parameter estimates, optimally weighting them to minimize variance of between-school estimates. Third, EB allows a fully multivariate formulation which adjusts all estimates for the interdependence among the multiple regression coefficients estimated within schools. Finally, EB is very flexible: sample sizes per school can vary greatly without biasing estimates, without throwing out data or "plugging in" guesses for missing values, and the investigator can make a variety of assumptions of error variances and covariances. The key restriction is that data within schools are assumed normal; and within-school paramters are assumed to have a normal "prior" distribution. Further the consequences of violating these assumptions are not yet well known.

The strengths of the EB approach are illustrated by a reanalysis of the High School and Beyond data.

The "Common School" Effect

Recently Coleman et al (1982) have inspired a re-examination of the "excellence vs. equity" issue in American High Schools. They found that Catholic high schools promote higher achievement than public high schools and that this positive effect of Catholic schools was most pronounced for lower SES students. We show how EB methods can be employed systematically to investigate this assertion. We begin with a simple two-stage model with SES as a predictor within schools and no predictors between schools.

We then demonstrate how to add predictors at each level until a model is discovered which adequately accounts for variation at both levels.

Model I

Within school i (i=1,2,...k=176) mathematics achievement varies as a function of student social class and random error:

$$y_{it} = \mu_i + \beta_i(x_{it} - x_{i.}) + R_{it}$$

y_{it} = mathematics achievement for student t in school i;

μ_i = the mean math achievement for school i;

β_i = the effect of SES on math achievement within school i;

x_{it} = the SES of student t in school i;

$x_{i.}$ = the mean SES for school i;

R_{it} = the error of estimate for student t in school i. Thus a large error indicates that knowledge of that student's SES was unhelpful in estimating the student's math score.

Between schools, we assume first that all variation of within-school parameters (i.e., all means and slopes) is random variation around a grand mean:

$$\mu_i = \mu + U_{0i}$$

$$\beta_i = \beta + U_{1i}$$

μ = the grand mean achievement across all schools;

U_{0i} = the effect of school i on mean math achievement;

β = the average effect of SES on achievement pooled within all schools;

U_{2i} = the effect of school i on the SES/math relationship;

By employing the EM algorithm (Dempster, Laird, and Rubin, 1977) it is possible to estimate the proportion of variance of estimated school means and slopes which represents parameter variance as opposed to variance of errors of estimation. Clearly this proportion has important implications for further study. For instance if virtually all variance of slope estimates were attributable to error, we would infer that schools are very much alike in the strength of the SES/math relationship. It would then be of little use to hunt for school policies and practices which "explain" such variance. On the other hand if a substantial portion of variance is parameter variance, it makes good sense to search for such explanatory variables. Re-estimation of these variance components then enables a reassessment of model adequacy.

Model II

Next we again estimate the same model as before "within schools," but now add a predictor "between schools:" school sector (0 = public, 1 = Catholic).

The second stage of the model now becomes

$$\mu_i = \gamma_{00} + \gamma_{01} w_i + U_{0i}$$

$$\beta_i = \gamma_{10} + \gamma_{11} w_i + U_{1i}$$

γ_{00} = the mean achievement for public schools;

γ_{01} = the "sector effect;" the mean difference between Catholic and public schools on mean math achievement

U_{0i} = The discrepancy between school i's mean and the mean for school i's sector.

γ_{10} = the average effect of SES on math achievement within public shcools;

γ_{11} = The mean difference between public and Catholic schools on the strength of the SES effect within schools.

w_i = sector: 0 if a school is public; 1 if Catholic.

U_{1i} = the discrepancy between the effect of SES on achievement within school i and the average effect for school i's sector.

Model III

We now add a "within-school" predictor: hours of homework per week. We add this factor because we hypothesize that it may attenuate the effect of SES within schools. The model between schools remains the same: with a single predictor, school sector. The within-school model now becomes:

$$(3.5) \qquad y_{it} = \mu_i + \beta_{1i}(x_{1it} - x_{1i.}) + \beta_{2i}(x_{2it} - x_{2i.}) + R_{it}$$

β_{1i} = effect of SES within school i;

x_{1it} = the SES of student t in school i;

β_{2i} = the effect homework within school i.

x_{2it} = hours per week of homework done by student t in school i.

Model IV

We now add two predictors between schools: the mean SES of the school and the mean SES by sector interaction effect:

$$(3.7a) \qquad \mu_i = \gamma_{00} + \gamma_{01}w_{1i} + \gamma_{02}w_{2i} + \gamma_{03}w_{1i}w_{2i} + U_{0i}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}w_{1i} + \gamma_{12}w_{2i} + \gamma_{13}w_{1i}w_{2i} + U_{1i}$$

$$\beta_{2i} = \gamma_{20} + \gamma_{21}w_{1i} + \gamma_{22}w_{2i} + \gamma_{23}w_{1i}w_{2i} + U_{2i}$$

Here w_{2i} is the mean SES of the students in school i, and γ_{02}, γ_{12}, and γ_{22} are the effects of mean SES on mean math achievement, the SES/Math achievement slope, and the Homework/math slope, respectively. Also, γ_{03}, γ_{13}, and γ_{23}, are the effects of the mean SES-by-sector interaction.

We stop here. Our purpose is to illustrate the use of the method for controlling variation within and between schools, not to find the optimal model.

Results

The results of estimating the four models are summarized in Table 1. Key findings are the following.

Model I

On average, there is an unmistakable linear relationship between student SES and mathematics achievement within schools, a result which is hardly surprising. There is, however, substantial variation among the schools in this effect, after removing that part of the variation among the estimated slopes solely attributable to their unreliability of estimation. In fact, about 35% of the total variance of the estimated slopes is estimated to reflect variation among the parameters.

The relative unreliability of the slopes as outcomes is illustrated, however, by the fact that, in contrast, 92% of the variance of the school means is estimated to be systematic, that is, to be variance of the parameters.

Model II

Catholic schools are found to have 1) substantially higher mean achievement than public schools; and 2) substantially smaller slopes, illustrating the egalitarian effect found by Coleman and his associates. These results are illustrated graphically in Figure 1.

Inclusion of sector accounts for 71.6% of the original estimated variation of the slope parameters, but because of the unreliability of these estimates, only 25% of the total variance. Inclusion of sector is less helpful in explaining variation in school means: 11.3% of the parameter variance and 10.2% of the total variance is explained.

Model III

Inclusion of homework within schools 1) helps very modestly in explaining within-school variation; and 2) leads to a very small adjustment of the effect of sector on slopes. The egalitarian effect of Catholic schools remains largely intact.

Model IV

Inclusion of the mean SES and the mean SES-by-sector interaction has several important effects on estimates:

1. The Catholic school advantage in mean math achievement disappears.

2. The "egalitarian effect" of Catholic schools remains intact: SES effects within Catholic schools remain substantially smaller, on average, than those within public schools.

3. Combining evidence from 1) and 2) yields the inference of a disordinal interaction between school sector and pupils' SES; Catholic schools appear to benefit poorer students but to penalize more advantaged students. This inference, like others, is quite tentative since we might plausibly revise these estimates in light of new information yielded by more complex models.

4. A substantial proportion of the variance in the slope parameters has been explained: 83.2%. (Only 29.0% of the total slope variance --which includes error variance -- has been explained.) The comparable figures for school means are less substantial but still impressive: 66.2% and 60.0%.

It is now possible, in fact, to retain the null hypothesis that all variance among the slope paramters has been explained. We employ the fit statistic proposed by Hedges (1982):

$$\sum 1/v_i(\hat{\beta}_{1i} - \gamma_{10}^* - \gamma_{11}^* W_{1i} - \gamma_{12}^* W_{2i} - \gamma_{13}^* W_{1i} W_{2i})^2$$

where v_i denotes the sampling variance of the SES/math slope for school i. If we assume that these estimated variances are equivalent to their true values, this statistic has a chi-squared distribution with k-4 = 172 degrees of freedom. In this case the statistic has a value of 193.07, which is equivalent to a unit normal deviate of 1.13.

Conclusions

This paper has illustrated a systematic approach for adjusting simultaneously for effects measured ar two levels of aggregation. It has also shown how to assess model adequacy in each stage of an iterative process whereby variables "within" and "between" are added. This strategy, known as empirical Bayes estimation,has obvious attractions for educational researchers who commonly confront multi-level data which has proved resistant to satisfactory quantitative assessment. We showed, for instance, how to resolve the regression of math achievement on SES into three components: a between-school component; a pooled within-school component; and a school-specific component. Further, the estimates of each of these components was adjusted simultaneously for potentially confounding variables within and between schools. This simultaneous, multi-level strategy offers many opportunities for gains in educational research.

Analysts are advised, however, to consider the tenability of key assumptions: that within-school data are normally distributed, and that the parameters whose variance is to be explained are normally distributed. Under these conditions the estimates used here are maximum likelihood estimates with concomitant advantages of asymptotic efficiency and known asymptotic distributions. Little is yet known about the consequences of violating these assumptions.

Table 1.

Summary of Results

FIXED EFFECTS	Model 1 (basic) effect	s.e.	Model 2 (SECTOR @ stage 2) effect	s.e.	Model 3 (HMWORK @ stage 1) effect	s.e.	Model 4 (SchSES + int. @ stage 1) effect	s.e.
Overall Effects								
grand mean MATHACT, γ^*_{00}	100.74	.573						
mean SES/MATHACH relation., γ^*_{10}	4.52	.290						
mean HMWORK/MATHACH rel., γ_{20}								
Sector Effects								
mean MATHACH, γ^*_{01}			5.06	1.09	5.06	1.09	.43	.83
SES/MATHACH relation., γ^*_{11}			-3.86	.49	-3.56	.49	-4.24	.45
HMWORK/MATHACH relation., γ^*_{21}					-.33	.10	-.36	.11
School SES Effects								
mean MATHACH, γ^*_{02}							15.09	1.32
SES/MATHACH relation., γ^*_{12}							2.32	.90
HMWORK/MATHACH relation., γ^*_{22}							.11	.19
School SES x SECTOR Effects								
mean MATHACH, γ^*_{03}							-2.50	1.85
SES/MATHACH relation., γ^*_{13}							.41	1.28
HMWORK/MATHACH relation., γ^*_{23}							-.39	.25
RANDOM EFFECTS								
variation among school means in MATHACH								
est. var $(\mu_i \mid \omega_i)$		53.25		47.36		47.24		18.01
est. var $(\hat{\mu}_i \mid \omega_i)$		58.66						
			$R^2 = 0.102$		$R^2 = 0.102$		$R^2 = 0.600$	
% of variance systematic		0.92	$R^2_s = 0.113$		$R^2_s = 0.113$		$R^2_s = 0.662$	
variation among SES/MATHACH slopes across schools								
est. var $(\beta_{1i} \mid \omega_i)$		5.26		1.49		1.48		0.88
est. var $(\hat{\beta}_{1i} \mid \omega_i)$		15.11						
			$R^2 = 0.250$		$R^2 = 0.252$		$R^2 = 0.290$	
% of variance systematic		0.35	$R^2_s = 0.716$		$R^2_s = 0.719$		$R^2_s = 0.832$	

Note, numerous other variances and covariances among the parameters could be estimated depending upon the substantive problem of interest.

Figure 1.

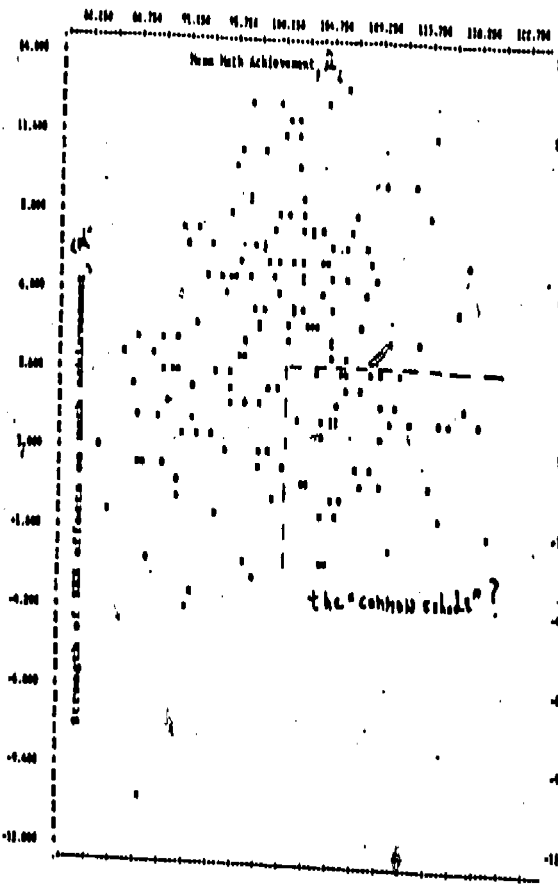

Figure 3.1: Mean math achievement (horizontal axis), and strength of the effect of student SES on math achievement (vertical axis) within 176 US high schools.

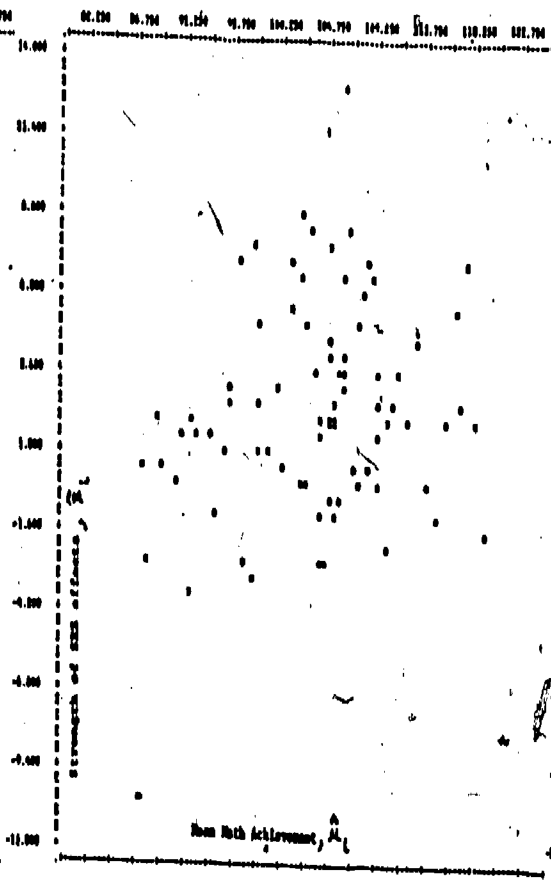

Figure 3.2a: Mean math achievement (horizontal axis) and effect of SES on math achievement (vertical axis) within 82 US Catholic schools.

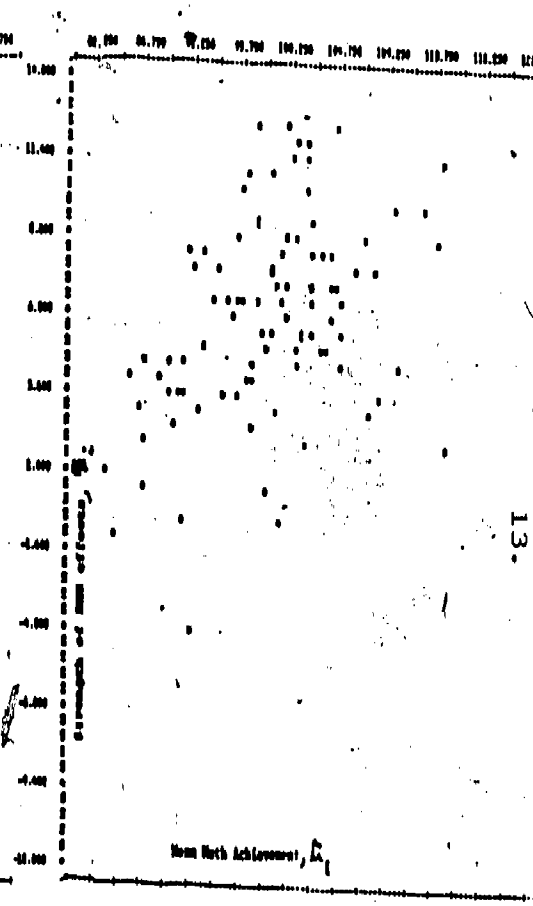

Figure 3.2b: Mean math achievement (horizontal axis) and effect of SES (vertical axis) within 94 US public high schools.

13

Note and References

Note 1. The results reported in this paper appeared in Chapter 3 of Raudenbush's (1984) doctoral dissertation.

Burstein, L., and Miller, M.D., Alternative analytic models for identifying educational effects: where are we? Paper presented and the annual meeting of the American Educational Research Association, Toronto, Ontario, 1978.

Coleman, J.S., Hoffer, T.E., and Kilgore, S. High School Achievement: Public, Private, and Catholic Schools Compared. 1982, New York: Basic Books.

Cooley, W.W., Bond, L., and Mao, B. Analyzing multi-level data. In R.A. Berk (ed.) Educational Evaluation Methodology. 1981, Baltimore: Johns Hopkins University.

Dempster, A.P., Laird, N.M., and Rubin, D.B., Maximum likelihood from incomplete data via the EM algorithm (with discussion), Journal of the Royal Statistical Society Series B, 39,1977,1-8.

Hedges, L.V. Fitting continuous models to effect size data. Journal of Educational Statistics. Winter1982, 7, 4, 245-270.

Lindley, D.V., and Smith, A.F.M., Bayes estimates for the linear model. Journal of the Royal Statistical Society Series B, 34, 1972, 1-41.

Raudenbush, S. W., Applications of a Hierarchcial Linear Model in Educational Research, unpublished doctoral dissertation, Harvard University Graduate School of Education, January, 1984.

Strenio, J.L.F., Empirical Bayes estimation for a hierarchcial linear model. Unpublished doctoral dissertation, Harvard University Department of Statistics, 1981.

Strenio, J.L.F., Weisberg, H.I., and Bryk, A.S., Empirical Bayes estimation of individual growth curve paramters and their relationship to covariates, Biometrics, 39, 1983, 71-86.